常见图论数据构造方法的研究

葛昌威

（重庆交通大学信息科学与工程学院，重庆 404100）

摘要 本文针对算法竞赛与数据结构课程中常见的图论模型，提出了兼顾随机性与时间效率的构造算法，并对部分构造方法的时间复杂度进行了证明。

关键词 图论；算法；调和级数；几何分布

A Study of Common Graph Theory Data Construction Methods

Ge Changwei

(Information Science and Engineering,Chongqin jiaotong University,Chongqin,China,404100)

Abstract In this paper, we propose construction algorithms that balance randomness and time efficiency for graph theoretic models commonly found in algorithmic contests and data structure courses, and prove the time complexity of each construction method.

Keywords Graph Theory，Algorithm，Harmonic series，Geometric distribution

1 概述

在算法竞赛与涉及图论的课程中常见的图大体上可以分为树，有向无环图，有向带环图，负环图，二分图，仙人掌等，本文将按边权的构造，一般图的构造，特殊图的构造依次进行介绍。

2 边权的构造

构造图的边权等价于构造一个数列，该数列需要满足以下四个数据限制:

M:数列长度。

L:数值下界。

R:数值上界。

unique:数列中数值是否可重复。

对于数值可重复的序列的构造方法较为简单，使用随机数构造即可。

而对于数值不可重复的序列的构造，需要根据 M，L，R 的取值进行分类讨论:

法 1:对于 R−L 较大的情况，我们可以采用 std::rand()取随机数，再用 std::unordered_map 对重复值进行去重。

法 2:对于 R-L 较小的情况，可以直接将[L,R]区间中所有数值存入数组，随机打乱数值后取前 M 项。

在[L,R]区间中进行若干次随机取值，取到 M 个不同的数字的期望次数。这个问题的模型实际上是由 M 个几何分布的模型构成，如果我们已经取到了 X 个不同的数字，设 N=R-L+1 那么下一次想要取得一个未出现的数字的成功率为:

$$\frac{N-X}{N} \qquad (1.1)$$

在一个几何分布中，每次实验的成功率为 P，则其期望次数为 P 的倒数。

由此我们可以得到在[L,R]区间中进行若干次随机取值，取到 M 个不同的数字的期望次数为[2]:

$$\sum_{i=1}^{M}\frac{N}{N-i+1} \quad (1.2)$$

使用 std:: unordered_map 进行重复值判断的均摊时间复杂度为$O(1)$。将整个数值区间打乱后取前 M 项的时间复杂度为$O(N)$。

根据实际复杂度进行判断。若以下表达式为真，则采用法 1，反之则采用法 2。

$$\sum_{i=1}^{M}\frac{N}{N-i+1} \leq N \quad (1.3)$$

$$\sum_{i=1}^{M}\frac{1}{N-i+1} \leq 1 \quad (1.4)$$

显然，计算这个表达式的时间复杂度是$O(M)$的，我可以利用近似值来优化时间复杂度。

当 M=N 时，可以发现(1.4)左式为调和级数，而调和级数有不等式[3]:

$$ln(N+1) \leq \sum_{i=1}^{N}\frac{1}{N-i+1} \leq (ln\ N)+1 \quad (1.5)$$

(1.6)

$$\sum_{i=1}^{M}\frac{1}{N-i+1} = \sum_{i=1}^{N}\frac{1}{N-i+1} - \sum_{i=1}^{(N-M)}\frac{1}{(N-M)-i+1}$$

根据(1.5)，我们可以用求出(1.6)左式的一个近似解:

$$\sum_{i=1}^{M}\frac{1}{N-i+1} \approx (ln\ N+1) - ln((N-M)+1) \quad (1.7)$$

将(1,7)带入到(1,4)得到新的判别公式:

$$(ln\ N+1) \leqslant 1 + ln((N-M)+1) \quad (1.8)$$

使用 C++中的 std::log(),可以$O(1)$地进行判断。

3 简单有向图的构造

由于非连通图可以由多个连通图组合而成，故这里只讨论连通图的构造方法。

假设要构造一个包含 N 个节点，M 条边的图，先根据 N 个节点构造出一颗树。构造树的伪代码：

```
for i in range(2,N):
    u=rand(1,i-1)
    v=i
    add_edge(u,v)
```

将这颗树的节点进行拓扑排序，从拓扑序列中选取两个节点。只需保证出发点拓扑编号均小于结束点，即可保证图无环。而想要保证图有环，只需要将出发点与结束点调转即可。

简单图要求不能存在重边与自环，所以在构造边时仍需根据边密度进行构造。

若边密度较低，可以用随机数在拓扑序中选点，使用 map<pair<int,int>,bool>进行去重。

而如果密度较高，可用类似于序列构造的方法，先构造树，再枚举出剩下所有的边，打乱后选取前 M-N+1 条边。

使用何种构造方法可以参照(1.8)式进行判断。值得注意的是，有向无环图的最大边数为$\frac{n^*(n-1)}{2}$，而带环图的最大边数为n^2。

4 特殊图的构造

4.1 负环图的构造

如果图中至少存在一个环，并且这个环上的边权总和为负数，这个图就被称为负环图。构造方法需要根据边权的取值范围和边的数量决定。

1:随机构造一个有向带环图，再对图中是否存在负环进行判断。

2:生成一个树，非随机地构造一些负环连接到树上。

该假定生成的负环图需要满足以下四个数据限制:

N:节点数量。

M:边数量。

L:数值下界(L<0)。

R:数值上界(R>0)。

很容易证明，一个有向带环图中的任意一个环，其为负环的概率为

$$\frac{|L|}{R+|L|}$$

如果我们采用第三节的方法来构造带环图，图中一共会包含 M-N 个环，则构造出的图中不包含负环的概率为:

$$\left(1-\frac{|L|}{R+|L|}\right)^{M-N}$$

对于|L|=R 的情况下，如果 M-N>=7，我们就有 99%以上的几率至少构造出一个负环。

构造完成后，可以使用最短路算法 spfa 进行检测。spfa 算法的特点就是每个点在被更新后会重新入队，也就是如果出现了一个负环，负环中的点之间不断更新答案，就会导致负环上的点不断入队再出队。

这里可以引入一个定理:入队次数超过节点总数的节点必定在负环上，在一个不包含负环的图中，对一个节点最多只能更新 N 次，也就是入队次数不会超过 N。当入队次数超过N就必然存在一个环导致其无限入队出队。

此外，我们可以将 spfa 中用于储存入队节点的队列换为栈，将新入队的节点放入栈顶，遇到环就会在栈顶连续出栈入栈。可以加速对负环的判断。

4.2 二分图的构造

节点由两个集合组成，且两个集合内部没有边的图被称为二分图。

最简单的构造方法是:将所有点随机分为两个集合，在两个集合之间随机选点连边。由此构造出的二分图无法保证连通性。

假设要构造有N个节点的二分图，首先构造一棵树，取任意节点作为根节点，根节点的深度为 0。其他节点的深度为父亲节点深度+1。

引理: 如果一个图中不存在奇环，那么该图就是二分图，否则该图不是二分图。

将所有节点按深度的奇偶性进行分组。从两组节点中各随机选一个节点进行连边，由于两点在原树上的距离为奇数，故生成的环必然为偶环。

4.3 仙人掌图的构造

如果一张无向连通图的每条边最多在一个环内，则称它是一棵仙人掌。

使用一棵树保证基本连通性，我们可以通过每个节点向父亲节点连一条边来实现构造一个边数尽可能多的仙人掌。但显然，这种构造方法缺少了随机性。

我们可以使用类似于树链剖分的方式来构造一个高随机性的仙人掌:

令top_x为节点 x 所在树链中，最接近根节点的节点编号。设son_x为节点 x 的子节点编号集合。对于$y_i \in son_x$，选定其中一个子节点，令 $top_{y_i}=top_x$，其他节点 $top_{y_i}=y_i$。

随后遍历所有点，如果:

1:top_x未被新增的边连接过。

2:$top_x \neq x$，并且top_x不是 x 的父亲节点。

那么就添加一条连接 x 与top_x的边。

树链中的任意一条边都只会被包含在一个环中。

按此种构造方法，最多可构造的环数 cnt 为:

$$\sum_{i=1}^{N}\left(size\left(son_i\right)-1\right)$$

使用第二节介绍的构造方法来构造树时，节点i会从$[1,i-1]$中选择一个节点 x 作为父亲。如果$x \neq i-1$，那么 cnt+1，否者 cnt 不变。故则最终 cnt 的期望值为:

$$\sum_{i=2}^{N}\frac{i-2}{i-1}$$

当 N 足够大的时候，我们可以近似地认为 cnt=N。但两者并不能判定为等号，因

为如果 M=2*N-2 时，只要有一个非根节点的子节点数量为 1，那么就无法构造。更为妥当的方法是，当 M>N*1.5 时，应采用低随机性的方法去构造。

参考文献

[1] OI-Wiki.图论相关概念[EB/OL]. https://oi-wiki.org/graph/concept/,2023

[2] M_Lter. LightOJ - 1248 Dice (III)(概率期望+几何分布)[EB/OL]. https://blog.csdn.net/lidengdengter/article/details/82527426, 2018

[3] Accelerator.1/1+1/2+1/3+1/4+……+1/n=?[EB/OL].https://www.zhihu.com/question/46263998/answer/2211767878,2021